R E P O R T  R E S U M E S

ED 002 916 24

AN INVESTIGATION OF DIFFERENTIAL BINAURAL STIMULATION IN THE TEACHING OF A FOREIGN LANGUAGE.

BY- VAN RIPER, CHARLES
WESTERN MICHIGAN UNIV., KALAMAZOO

REPORT NUMBER CRP-739 PUB DATE OCT 68
EDRS PRICE MF-$0.09 HC-$2.20 55P.

DESCRIPTORS- *SPANISH, COMPARATIVE ANALYSIS, LANGUAGE LABORATORY EQUIPMENT, LANGUAGE LABORATORY USE, *LANGUAGE INSTRUCTION, LANGUAGE TESTS, PRONUNCIATION, AURAL STIMULI, KALAMAZOO, MICHIGAN

THIS STUDY DETERMINED WHETHER OR NOT DIFFERENTIAL BINAURAL STIMULATIONS CAN BE USED EFFECTIVELY TO IMPROVE PRONUNCIATION IN FOREIGN LANGUAGE TEACHING. THE OBJECTIVE WAS TO DETERMINE WHAT EFFECT HEARING SIMULTANEOUSLY THE TEACHER'S VOICE IN ONE EAR AND HIS OWN VOICE IN THE OTHER WOULD HAVE ON A STUDENT'S ABILITY TO COMPARE THE DIFFERENCES IN PRONUNCIATION. TWO GROUPS OF FIRST-SEMESTER FRESHMAN COLLEGE STUDENTS WITHOUT PREVIOUS TRAINING IN SPANISH SERVED AS SUBJECTS. THEY RECEIVED GENERAL LANGUAGE-LABORATORY TRAINING. THE EXPERIMENTAL GROUP WAS GIVEN DIFFERENTIAL BINAURAL STIMULATION, HEARING THEMSELVES IN ONE EAR AS THEY HEAR THE TAPED STIMULUS MATERIAL IN THE OTHER. THE CONTROL GROUP WAS GIVEN THE SAME STIMULUS MATERIAL BUT HEARD IT IN BOTH EARS ALONG WITH THEIR OWN VOICES. TESTING MATERIALS WERE ADMINISTERED TO BOTH GROUPS AT THE END OF 3 CONSECUTIVE MONTHS OF TWICE-A-WEEK LANGUAGE-LABORATORY TRAINING. WITH PRONUNCIATION USED AS A BASIS, AN ANALYSIS OF VARIANCE AND THE STATISTICAL SIGNIFICANCE OF THE DIFFERENCES WERE COMPUTED. DIFFERENTIAL TRAINING APPEARED TO HAVE BEEN INEFFECTIVE. NO STATISTICALLY SIGNIFICANT DIFFERENCES WERE FOUND BETWEEN EXPERIMENTAL AND CONTROL GROUPS IN NUMBER OF ERRORS OR IN RATING SCALE VALUES. (GC)

ED002916

OFFICIAL RESEARCH FILE COPY

Cooperative Research Project No. 739

# An Investigation of Differential Binaural Stimulation in the Teaching of Foreign Languages

ED002916

U. S. DEPARTMENT OF HEALTH, EDUCATION AND WELFARE
Office of Education
This document has been reproduced exactly as received from the person or organization originating it. Points of view or opinions stated do not necessarily represent official Office of Education position or policy.

COOPERATIVE RESEARCH PROJECT

# AN INVESTIGATION OF DIFFERENTIAL BINAURAL STIMULATION IN THE TEACHING OF A FOREIGN LANGUAGE

Charles Van Riper

(Principal Investigator)

Western Michigan University

Kalamazoo, Michigan

October, 1960

The research reported herein was performed pursuant to a contract with the United States Office of Education, Department of Health, Education and Welfare.

3414A
224

ACKNOWLEDGEMENT

Any project of this type requires the joint efforts of many individuals.

Listed below are the names of the staff and consultants who participated.

Charles Van Riper, Ph.D., Principal Investigator, Director, Speech Clinic, Western Michigan University

Herbert B. Jones, M.A., Asst. Prof. of Languages, Western Michigan University

Gustavo A. Fierro, Instructor, Western Michigan University

Marilyn Lamond, Ph.D., Asst. Prof. of Languages, Western Michigan University

Jan Verhagen, M.A., Research Associate, Western Michigan University

Louis Acuna, Native Spanish Speaker

Marcelle Dale, Language Dept., Kalamazoo College

Helen Grable, Language Dept., Central High School, Kalamazoo, Michigan

TABLE OF CONTENTS


I. INTRODUCTION AND BACKGROUND ..... 1

II. PROCEDURES ..... 8

III. RESULTS OF DICHOTIC VERSUS BINAURAL STIMULATION ... 12

IV. RELIABILITY AND VALIDITY OF EXPERT JUDGMENTS ..... 18

V. DISCUSSION ..... 24

VI. SUMMARY AND CONCLUSIONS ..... 27

VII. BIBLIOGRAPHY ..... 32a

VIII. APPENDIX A ..... 33

IX. APPENDIX B ..... 39

X. APPENDIX C ..... 41

XI. APPENDIX D ..... 44

## LIST OF TABLES


TABLE I. Results of dichotic versus binaural stimulation training.....12

TABLE II. Results of first post-training testing.....14

TABLE III. Results of final post-training testing.....16

TABLE IV. Reliability correlation of first and second judgments.....20

TABLE V. Agreement among judges.....22

# INTRODUCTION AND BACKGROUND

The increasing urgency for adequate communication between our citizens and those of other nations has spurred interest in methods for improving the teaching of foreign languages in our schools and universities. An evaluation of the current attitudes of language teachers themselves by Borglum (1) is far from optomistic however.

> "Our profession's problem today stems from the fact that the teaching of foreign languages as languages has become important. At the same time the training of teachers of language is neither about language, about teaching nor in any way scientific. On the contrary, it is literary and historical research, even at the undergraduate level. As a profession we language teachers have been indoctrinated with the ideal that our sole mission in life is to kindle the small bright blue flame in the worthy few. Anything which relates to the masses, to mass psychology, to mass media, to the common need, fails to touch our sense of responsibility or to arouse our interest, let alone our enthusiasm..."

> "Foreign languages are mostly taught as an academic exercise, directed from a chair located behind a table which is the teacher's fortress...."

Among the essential problems confronting the teacher of modern languages is the improvement in pronunciation skills. It is not enough for our students to have a reading knowledge of a language or to be able to speak it with so many distortions that comprehension by any native would be difficult.

The American who could serve our national interest abroad best is one who not only speaks the foreign tongue fluently but also with the accent, melody and phonemic characteristics of that language.

After pointing out that the Government has had to provide intensive training in foreign languages to prepare Foreign Service personnel, Johnston (5) asks:

> "1. Why do costly, last-chance, emergency training programs in Government have to be continued and intensified in the languages commonly taught in schools and colleges throughout the country?"
> "2. How have we defined 'useful' knowledge in foreign languages in our school programs? Do we yet have a clear statement about what degree of accomplishment can reasonably be expected at any given age in any given length of time?...."
> "3. If we begin to review facts regarding the present status of foreign language study, the salient reality is that few Americans learn to communicate in any language other than English and that even as our destiny becomes more intricately enmeshed with that of other peoples we are not, on any grand scale, doing the best we know how to do in the teaching of foreign languages."

This nation has fallen far behind its rival and friendly countries in the world in devising a system of education which will provide an adequate number of bilingual citizens to fulfill the requirements of science, politics, and other functions of interdependence. Our national survival itself may depend upon the swiftness with which we can remedy this situation.

Despite the advances made in the language schools set up in the armed services during the war, we still are woefully inefficient and primitive in our methods for teaching a foreign language. Moreover, we cannot assume that the kind of motivation which existed in the language schools during the war is present in our peace-time schools and colleges. We need a supply of language teachers but it is doubtful that we will be able to get them, or that they will be able to solve our need for bilingual man-power even if they are procured, because our present methods for teaching a foreign language leave much to be desired.

One approach to the problem of improving pronunciation of a foreign language is that of the development of the language laboratory. This has been described by Johnston and Seerley (6) as follows:

> "The development of the language laboratory presupposes that learning to speak and understand the language is an important objective of modern foreign language study. Although variously expressed, two assumptions underlie all the statements relating to the laboratory's purpose: (1) that systematic aural-oral practice is indispensable in learning to speak a modern foreign language, and (2) that the conventional classroom does not provide adequately for such practice.
>
> "The basic purpose of the language laboratory, therefore, is to provide regular practice in listening to good models of the foreign speech and a large amount of imitation and a large amount of imitation and repetitive oral drills...."

The advantages of such laboratory techniques are several. First it is economical of staff time; secondly, it insures that students get sufficient exposure to the foreign language; finally, it permits the use of materials prepared by native speakers.

While there are various methods for presenting the stimulus materials in language laboratory experience, the typical procedure is that in which the student dons headphones, listens to words or sentences and then repeats them after the model. The instructor may listen in on a monitor circuit or may merely stand behind the student as he repeats what he has heard. Often even this monitoring is absent, and the student is left to his own devices to improve his pronunciation.

Among the obvious criticisms of this procedure is that the headphones tend to prevent the student from hearing himself. When the student speaks in unison with the model, hearing that model in both ears, the ensuing masking is bound to create interference with the discrimination of error or the correction of response. Some teachers have attempted to remedy this procedure by feeding in the stimulus from the tape into one ear only. However, the relative intensities of the students own voice and that being amplified in his ear via the

headphone are far from equal. Other attempts have led to the recording of the student's speech during the laboratory experience and then analyzing the errors in a post-stimulation session. This has the advantage of objectifying the student's speech but it still does not present any opportunity for the point-to-point comparison which seems most effective in the recognition of error. In most of the sense modalities, simultaneous comparisons are more effective than successive ones in the development of discrimination. When the student of a foreign language hears a speaker's voice on tape or disc recording or hears his teacher and then repeats what he has heard, he must compare two memories if he is to analyze any errors which might occur. It was felt that some improvement in procedure might be gained if binaural stimulation were employed, the student hearing his own voice in one ear and hearing the voice of the stimulus tape in the other ear. This would provide simultaneous rather than successive comparison, and, theoretically, could facilitate the acquisition of adequate pronunciation skills.

Under such a differential binaural feedback system, it was hypothesized that the student would no longer have to compare two memories after delay. By hearing both his own and the stimulus speech at the same time, it was possible that

discrepancies and errors would become more vivid and might then be eliminated more readily. Since such subtle skills as stress, duration, inflection demand keen discrimination, it seemed possible that, by matching his own speech simultaneously with that of the taped model, these features of deviant articulation and utterance could be more readily eliminated.

Dichotic stimulation (the separate stimulation of the two ears) has played a part in psychological experiments for many years, usually being employed in sound localization experiments. We know that simultaneous tones of the same frequency and composition are fused perceptually to result in an apparent single tone. Two pure tones must differ by as much as 6 to 16 cycles before they can be separated and assigned to the ear stimulated. Two complex tones, of the same pitch, can easily be discriminated when presented dichotically. Recently, binaural and dichotic stimulation has been receiving some attention by audiologists. Koenig (7) found that monaural hearing of several simultaneous conversations was impossible. In contrast, binaural hearing permitted the identification and separation of the various speakers and "it was fairly easy to understand speech under conditions of extremely high room noise." Carhart (3) and others have

stressed the usefulness of the binaural hearing aid, pointing out that it improves not only sound localization but speech recognition. Schubert (10) measured the intelligibility of speech under masking noise in three binaural listening conditions: 1. simultaneously and in phase; 2. simultaneously but in opposite phase; and 3. speech arriving at one ear later than the other. Conditions two and three gave about the same intelligibility and were better than condition one. Calearo (2) asked his subjects to repeat bisyllabic words which had been presented binaurally so that one ear received a signal of normal quality and insufficient intensity while the other ear received a signal of normal intensity distorted in quality. Normally hearing subjects summated the two dichotic stimulations; patients with temporal lobe lesions did not. Licklider (8) has studied the effect upon speech intelligibility of interaural phase relationships. Cherry (4) has reported the difficulties inherent in speech recognition of different messages presented dichotically. Ptacek (9) presented pairs of stimulus words dichotically to his subjects simultaneously by means of stereophonic tape. They chose one of these words. He found that his subjects chose two-syllable in preference to one-syllable words, and the more familiar word in preference to the less familiar one. Van Riper (11)

has described and recommended the use of dichotic stimulation in speech therapy for the improvement of defective articulation. If this summary of the pertinent research appears rather impoverished it is because dichotic stimulation as an aid in teaching or therapy has received little attention.

## PROCEDURES

In order to determine whether or not dichotic stimulation, using the simultaneous hearing of the instructor's voice in one ear and the subject's own voice in the other, was more effective than binaural hearing of the instructor's voice in both ears in the learning of foreign language pronunciation, the following experiment was designed. Two groups of subjects were selected from students enrolling in the first course in Spanish at Western Michigan University. There were 53 subjects in the entire experimental group and 52 in the control group, a total of 105. They were matched in terms of High School academic achievement, previous foreign language training, travel in foreign countries, age and sex. Students with hearing or speech defects were rejected and each student and his control were enrolled under the same instructor for general language instruction, but those in the experimental group were given dichotic stimulation while those in the control group were given binaural stimulation in their language laboratory training. After two weeks of general language instruction, an initial test of pronunciation was administered to all students in both experimental and control groups. In this, each subject tape-recorded the words

and sentences he heard binaurally through the headphones. These taped samples were then evaluated by three judges, instructors in Spanish at the University. Each judge worked independently, encircling from the test sheets those words on which pronunciation error had occurred. Each taped sample was also evaluated in terms of a five step scale of pronunciation adequacy.

Then 23 training sessions were administered to all subjects, using a series of 7 training tapes. The experimental group heard these training tapes monaurally in their preferred ear; the controls heard them binaurally. Each single word or sentence in the training tapes was presented four times. The subjects were to say the word (or the sentence) in unison with the model on the third presentation only. The subjects in the experimental group heard their own voices in one ear during the third stimulus presentation; those in the control group heard the voice of the model in both ears as they spoke the stimulus word on the third presentation. Transcripts of the training tapes will be found in Appendix A.

At the end of this initial training period of 23 sessions, the pronunciation test was administered to the 46 experimental and 42 control subjects who were still enrolled in Beginning Spanish and the same three judges again evaluated the

recorded samples.

Then, at the beginning of the second semester, a new series of training tapes (Appendix B) were prepared and administered, this time to a smaller number of the same subjects, the usual processes of academic attrition having reduced those enrolling. There were 33 subjects in each of the two final training groups and they were subjects who had previously been assigned throughout the first semester to the same experimental or control groups. A new and longer test of pronunciation ability was devised and this was administered and scored by the same judges at the end of the second semester after 22 additional training sessions. Each subject therefore was given a total of 45 training sessions either in dichotic (experimental) or binaural (control) stimulation. The results of all tests were subjected to an analysis of variance and the effects of the differential training were ascertained.

In order to determine the reliability of the evaluations made by the judges, 30 of the initial test samples were re-evaluated by the same three judges after an interval of three months. The order of sample presentation was varied. Correlations of first and second judgments of these same samples were computed. In addition, to determine the validity of the

judgment, three additional judges (also professional teachers of Spanish) were asked to judge these same thirty taped samples and the agreement between both groups of judges was found.

The results of the testing were expressed in terms of (1) number of words upon which all judges agreed that pronunciation error had occurred, (2) total number of words on which at least one judge had indicated error and (3) the average of the three scale values scored for each sample by the judges.

All tapes, both training and testing, were dictated by native Spanish speakers, upon high fidelity Scotch magnetic tape, using a Magnecord tape recorder, Model S-36-B. They were played back to the students in the language laboratory from the same Magnecord tape recorder through fifteen listening units. These units were installed in a battery of 15 sound treated booths. Each booth contained a Model SAU-10 Stromberg Carlson amplifier, a Model 70-14 Trimm binaural headphone set, a Model 924 Electro-Voice Lavalier microphone and then necessary switches and connecting cables. A supervisor was in constant attendance during each testing and training session to insure proper response.

## RESULTS

The results of this investigation fall under two headings: those bearing the results of differential training in pronunciation and those dealing with the validity and reliability of expert judgments of pronunciation adequacy.

### RESULTS OF DICHOTIC VERSUS BINAURAL STIMULATION TRAINING

TABLE I

| | Word Errors Checked by Three Judges | | | Total Error Words Checked by at Least One Judge | | | Average Scale Value of Judges ratings | | |
|---|---|---|---|---|---|---|---|---|---|
| | Mean | S.D. | t | Mean | S.D. | t | Mean | S.D. | t |
| Experimental Group (N = 53) | 8.16 | 3.01 | | 12.26 | 6.10 | | 5.20 | 3.80 | |
| | | | 0.40 | | | 0.54 | | | 0.22 |
| Control Group (N = 52) | 8.42 | 3.62 | | 13.32 | 7.22 | | 5.36 | 3.47 | |

As the results in Table I indicate, at the beginning of the training period, both experimental and control groups were judged as equal in terms of number of words on which all three judges agreed errors had occurred. The mean number of these word errors for the experimental group was 8.16 with a standard deviation of 3.01, and the mean and standard deviation of the word errors for the control group were 8.42 and 3.62

respectively. The t value was 0.40, which indicates a complete lack of difference between the groups since, with 105 degrees of freedom, the t value must be at least 2.63 to be statistically significant at the one per cent level of confidence, and at least 1.98 at the five per cent level of confidence. So far as word errors on which all the judges agreed, the two groups were alike at the beginning of the training period.

When the two groups were compared on the basis of all different words checked by at least one judge as having been mispronounced, the same lack of statistical difference was manifested. For the experimental group the mean number of all words so checked was 12.26 with a standard deviation of 6.10 and the mean of the control group was 13.32 with a standard deviation of 7.22. The t value for this comparison was also very low: 0.54 and we may conclude that on this measure the two groups were also equal at the beginning of training.

When the judges' ratings on the seven point scale of pronunciation adequacy were averaged and these scores submitted to an analysis of variance, the experimental group showed a mean scale value of 5.20 with a standard deviation of 3.80 while the controls had a mean scale value of 5.36 with

a standard deviation of 3.47. The t value for this comparison was 0.22, indicating no difference existed between the groups so far as ratings of pronunciation adequacy were concerned. At the beginning of the training period, then, on all three measures of pronunciation ability, the two groups were equal.

In Table II the results of the first post-training test are given. This test followed 23 training sessions and was identical with that given at the beginning of the experiment. This test is transcribed in Appendix D. As the table indicates, this first period of training in which the experimental group received dichotic and the controls received binaural stimulation showed no effect so far as pronunciation adequacy was concerned.

RESULTS OF FIRST POST-TRAINING TESTING

TABLE II

| | Word Errors Checked by Three Judges | | | Total Error Words Checked by at Least One Judge | | |
|---|---|---|---|---|---|---|
| | Mean | S.D. | t | Mean | S.D. | t |
| Experimental Group (N = 46) | 5.43 | 2.62 | 0.61 | 10.18 | 4.72 | 0.09 |
| Control Group | 5.09 | 2.57 | | 10.47 | 5.03 | |

When the groups are compared on the basis of word errors upon which all judges agreed, the mean of the experimental group was 5.43 with a standard deviation of 2.62 and the mean of the control group was 5.09 with a standard deviation of 2.57. The t value for this comparison was 0.61 which is not significant. For 82 degrees of freedom (there were 42 subjects in each of the two groups) t must be greater than 2.64 to be significant at the one per cent level of confidence and greater than 1.66 at the five per cent level. So far as words upon all three judges agreed errors had occurred, both groups were still equal.

When the two groups were scored on the basis of all words upon which at least one judge had indicated an error had occurred, the same lack of a statistically significant difference appears. The mean number of these total word errors for the experimental group was 10.18 with a standard deviation of 4.72. For the controls, the mean was 10.47 with a standard deviation of 5.03. The t value for this comparison was only 0.09, indicating that no difference existed between the groups.

When the rating scale values of the two groups were compared we find again no statistically significant difference. The mean scale value for the experimental group was 4.56 with a standard deviation of 1.58. For the control group the mean

scale value was 5.01 with a standard deviation of 1.71. The t value was 1.32 which is not significant at either the one or five percent levels of confidence.

We must conclude, then, that the initial period of 23 training sessions in differential stimulation had no effect upon pronunciation adequacy.

In Table III are given the results of the final testing, at the end of the second semester of the academic year, when 45 training sessions had been completed.

RESULTS OF FINAL POST-TRAINING TESTING

TABLE III

| | Word Errors Checked by Three Judges | | | Total Error Words Checked by at Least One Judge | | | Average Scale Value of Judges Ratings | | |
|---|---|---|---|---|---|---|---|---|---|
| | Mean | S.D. | t | Mean | S.D. | t | Mean | S.D. | t |
| Experimental Group (N=33) | 6.15 | 3.68 | | 14.10 | 7.68 | | 4.48 | 3.16 | |
| | | | 1.46 | | | 0.11 | | | 0.01 |
| Control Group (N=33) | 4.60 | 3.58 | | 13.72 | 6.20 | | 4.49 | 3.66 | |

It should be noted that this final testing did not use the same test employed at the beginning and at the end of the first training sessions. It was longer and, perhaps, more

carefully designed. In view of this change, it is not possible (nor necessary) to compare the results of the initial and final tests. We are primarily interested in determining whether or not the further training had effected an improvement in the pronunciation ability of the experimental group. It is obvious from an examination of Table III that this improvement did not occur. The two groups were still equal on all three measures.

In terms of word errors upon which all three judges agreed, the more flagrant errors, the experimental group showed a mean of 6.15 with a standard deviation of 3.68, while the control group had a mean of 4.60 with a standard deviation of 3.58. The t value is larger than those previously found, being 1.46 but it would need to be greater than 2.70 to be statistically significant at the one per cent level of confidence with 64 degrees of freedom and greater than 2.02 at the five per cent level. There is still no difference between the groups on this measure of pronunciation ability.

In terms of <u>total</u> words checked as being in error by at least one judge, the experimental group showed a mean of 14.10 with a standard deviation of 7.68 and the control group a mean of 13.72 with a standard deviation of 6.20.

The t value was only 0.11 which means that on this measure there was no statistically significant difference between the groups.

When the rating scale values are compared, again no difference between groups is shown at the conclusion of the year's training. The experimental group had a mean of 4.48 with a standard deviation of 3.16; the controls, a mean of 4.49 with a standard deviation of 3.66. The t value for this comparison was 0.01. There was no difference shown between groups in terms of reting scale values.

We must therefor conclude that, as judged and under the conditions of this experiment, the use of dichotic stimulation was no more effective in the improvement of pronunciation in Spanish than was the use of binaural stimulation.

## THE RELIABILITY AND VALIDITY OF EXPERT JUDGMENTS OF ADEQUACY IN SPANISH PRONUNCIATION

Among the aims of this study was the exploration of methods for measuring adequacy of pronunciation in a foreign language. There are, so far as we have been able to ascertain, no objective measures or ways of measuring such adequacy. Usually, students of foreign languages are graded by their instructors on pronunciation adequacy as well as vocabulary acquisition, grammar, syntax and the other aspects of language

mastery. How reliable and how valid are these grades? The present research provides some data bearing upon this question. In order to test the hypothesis that dichotic stimulation would produce more gains in pronunciation adequacy than would binaural stimulation, expert judgments were used. The judgments were those of college instructors in Spanish. They were provided with copies of the test material spoken by the subjects in repeating the words and sentences dictated on the test tapes by a native speaker. They listened to the subject's taped responses and encircled upon the test copies those words on which they discerned a pronunciation error. They also indicated by a check mark on a linear five step scale their over-all estimate of the subject's general pronunciation adequacy. These word counts and scale values provided the objective data of this part of the experiment.

Reliability. After a period of three months had elapsed, the same three judges who had made their initial judgments were given the first thirty taped test samples to score again. The samples were randomized. Again these judges encircled the words upon which pronunciation error was discerned and checked a scale value for each subject's responses. The consistency of the judgments is indicated in terms of the

product moment correlations given in Table IV and Table V.

TABLE IV

RELIABILITY CORRELATION OF FIRST AND SECOND JUDGMENTS

| | Number of words mispronounced | Rating Scale values |
|---|---|---|
| JUDGE A | .92 | .74 |
| JUDGE B | .79 | .50 |
| JUDGE C | .82 | .78 |
| ALL JUDGES CONCURRING | .83 | .79 |

These correlations indicate first that judges may vary widely in their consistency of judgments of pronunciation adequacy. Two of our judges showed a greater reliability than did the third. Secondly, it is apparent that the method of identifying and counting mispronounced words may be reliable enough for research in the field of language learning at least so far as pronunciation is concerned. It also indicates that general judgments of pronunciation adequacy, based upon scale values of taped samples such as those used in this study, may be very unreliable and therefore unsuitable for such research.

In the lower line of Table IV are given the correlations of first and second judgments in which all three judges

concurred that a pronunciation error had occurred on a given word. These errors were evidently the obvious and apparent ones and we might expect that the reliability coefficient would be higher. The results show, however, that it is possible for a single judge to be more consistent in his judgments than a group of judges in theirs. This may be due to the presence of a single judge whose judgments were not reliable. In any future research, care should be taken to use only those judges whose reliability of judgments has been determined.

Validity. The validity of judgments of pronunciation adequacy must ultimately depend upon the subjective criteria set up by native speakers of a foreign language. When is a deviation in pronunciation an error? In the absence of objective criteria, the only way we have of answering this question is in terms of the degree of agreement shown by expert speakers of that language. In this experiment we attempted to assess this validity in terms of the agreement shown among themselves by the three judges. The results are shown in the correlations of Table V.

AGREEMENT AMONG JUDGES

TABLE V

| Number of Words Mispronounced | | Rating Scale Values | |
|---|---|---|---|
| JUDGE A<br>JUDGE B | .81 | JUDGE A<br>JUDGE B | .81 |
| JUDGE B<br>JUDGE C | .83 | JUDGE B<br>JUDGE C | .69 |
| JUDGE A<br>JUDGE C | .78 | JUDGE A<br>JUDGE C | .72 |

It is apparent that the agreement between any of these judges is not very great and that it varies with the pairing of judges. It is also apparent that there was a higher degree of agreement in terms of words labelled as mispronounced than there was for the rating scale values. The degree of agreement as to what constitutes a pronunciation error, however, is hardly sufficient for valid research on the effectiveness of various types of language laboratory experience. This finding therefore may explain the negative results of the differential training our experimental subjects received. If the judges themselves could not agree on the number of

errors demonstrated by students on their recorded test samples, it is difficult to be sure that the differential training was effective or not.

Since it was also possible that our initial panel of judges was not composed of experts or that these judges were incompetent, we submitted the same test samples to a second panel of three judges who presumably also were experts. Selecting only those words, upon which all three of each panel agreed, as the measure which should show the greatest degree of agreement, we correlated the judgments of the first panel with those of the second. The resulting correlation coefficient was .89. This indicates that at least so far as the more obvious and flagrant errors were concerned, the two groups of judges were pretty much in agreement and that this measure is a fairly valid one. On total words scored as mispronounced ($r=.73$) and on rating scale values ($r=69$), these judges were no more in agreement with each other than were those of the first panel.

We may conclude, from this attempt to measure reliability and validity that though certain individual judges may be able to score word errors fairly reliably, they do not tend to agree with each other sufficiently (except upon those words in which the pronunciation error is especially flagrant)

for purposes of research on the effects of differential training on pronunciation adequacy.

## DISCUSSION

The completely negative results of this experiment in differential stimulation in the teaching of pronunciation in a foreign language would seem to merit some discussion. It is obvious that, under the conditions of this experiment, the use of dichotic presentation of the subject's own response simultaneously with that of the model failed to produce any gains. Several explanations suggest themselves. First, it is conceivable that the amount of dichotic stimulation in any one training period was not sufficient to provide the necessary impact. The subjects were given only one opportunity in every four stimulations to hear themselves in one ear at the same time they heard the model in the other. This may not have been sufficient. Secondly, the fact that the subjects in the experimental group heard the four stimulations in but one ear while those of the control group heard them in both may have produced a differential stimulus intensity in favor of the controls. This increased impact of the model for the controls may have equalled the gain due to dichotic stimulation on the part of the experimental group. Both groups may have made

equal gains for different reasons. Moreover, it is possible that the simultaneous presentation of the model in one ear with the subject's voice in the other may have created some masking of both or either. This would have reduced the effectiveness of the simultaneous comparison so far as the discrimination of error is concerned. Finally, as Ptacek (9) demonstrated in his research, familiar words seem to have more stimulus value than unfamiliar ones in dichotic stimulation. Our subjects may therefore have listened only to their own voices rather than to both their own and that of the model.

At least one positive finding, however, has emerged from this research. Several methods for measuring adequacy of pronunciation in a foreign language have been developed and one of them has been shown to be fairly reliable and valid, viz. the counting of words on which a panel of experts agree a pronunciation error has occurred. In view of the complete lack of reliable or valid tests of pronunciation ability, this contribution may be of some value.

The development of adequate methods for measuring pronunciation adequacy in a foreign language would seem to be urgently needed. Future research may well investigate further

the promising leads developed in this study. Also, the adequacy of methods used in the typical language laboratory can, as this study has demonstrated, be subjected to objective testing. Further research could, for example, measure the effectiveness of the interval in successive presentations of the model and subject's voices monaurally, binaurally or dichotically. The effectiveness of varying intensities of the model's voice should also be ascertained and the optimal number of stimulations prior to the subject's response should be determined. The present research may serve as a pilot study for the investigation of these problems.

## SUMMARY AND CONCLUSIONS

This study was concerned with two problems: (1) the reliability and validity of college instructors' evaluations of student adequacy in Spanish pronunciation and (2) the effectiveness of dichotic stimulation as opposed to binaural stimulation in language laboratory experience.

From university students electing the first course in Spanish, after preliminary screening to eliminate speech and hearing defects, 105 subjects were selected and divided into experimental and control groups. Pairs were matched in terms of sex, high school grade averages, other foreign language study, age and travel experience. There were 53 subjects in the initial experimental group and 52 in the initial control group. Each subject and his control were enrolled in the same class under the same instructor for general language instruction but, in the language laboratory, those in the experimental group were given dichotic stimulation while their controls were given binaural stimulation. This differential training continued for two semesters.

In order to ascertain the effectiveness of the initial matching, no language laboratory experience was given until two weeks after the beginning of the course. Then, samples of Spanish pronunciation in repeating words and sentences

were tape recorded from each subject. These recorded samples were then judged individually by a panel of three college instructors in Spanish by having them encircle all words on a typescript of the recordings on which error had been perceived. Each judge also provided for each subject a general estimate of adequacy in pronunciation in terms of a point on a five step rating scale. Judges made their judgments independently of each other.

The results of this initial testing demonstrated that there were no statistically significant differences between experimental and control groups at the beginning of the training insofar as the number of errors encircled or scale values provided by the three judges.

Then the language laboratory experience was initiated. In two twenty-minute periods weekly, all subjects heard and repeated samples of recorded Spanish speech, dictated on the master tape by a competent native speaker. On these training tapes, each word or sentance was spoken four times. In listening to the tape, all subject listened to the stimulus material during its first two presentations, spoke the material in unison with the model on the third presentation and listened again on the fourth presentation.

The subjects in the experimental group were fitted with binaural earphones. These were so wired that these subjects heard themselves in one earphone at the same time they heard the model in the other earphone during the third presentation. This dichotic arrangement permitted a simultaneous comparison of the subject's own utterance with that of the recorded native speaker. The controls heard all samples in both ears for all presentations.

At the end of 23 training sessions, all subjects were given a recorded pronunciation test. This test consisted of tape recorded isolated words and sentences also dictated by the same native speaker. Each subject heard binaurally the sample twice, then repeated it into a microphone leading to a tape recorder. In this way samples of the student's own Spanish pronunciation were procured for evaluation by the judges.

The results of this first evaluation demonstrated that the differential training had been ineffective. No statistically significant differences were found between experimental and control groups in terms of number of errors or in terms of rating scale values.

At the beginning of the second semester, 33 subjects in the experimental group with their 33 matched controls were available for further training. Both groups were given 22 additional language laboratory training sessions in dichotic or binaural stimulation. At the end of this period, another and longer pronunciation test was administered to all subjects. Again no statistically significant differences were found which could be attributed to the differential training. It is necessary to conclude that, under the conditions of the experiment, dichotic stimulation had no advantage in the teaching of Spanish pronunciation skills.

However, we were also interested in the degree to which the judgments of our panel of experts were valid and reliable. Accordingly, 30 of the initial test recordings were re-evaluated by the panel after a lapse of three months. The results indicate that the consistency of first and second judgments in terms of number of counted err r words and in terms of rating scale values was often surprisingly low, the correlation coefficients ranging from .50 to .92. Rating scale values were more inconsistent than number of error words counted. It would seem that if proficiency in pronunciation of a foreign language is important, college instructors in that language should be able to judge that proficiency with more consistency than was shown under the relatively objective conditions of this experiment.

Since it was possible that our panel contained judges of less than average ability in the judgment of pronunciation adequacy, we procured another panel of three presumably qualified experts and required them to judge the same thirty recorded samples evaluated by the first panel. In terms of total number of error words, and in terms of rating scale values, this second panel of judges also showed a marked inability to agree among themselves. The one measure upon which both panels of judges did show agreement was on the more flagrant errors. These were the words upon which all judges of each panel had agreed that a pronunciation error had occurred. The degree of agreement for the two panels for these flagrant errors is reflected in the correlation coefficient of .89. This is sufficiently high for the comparison of groups in language training experiments, and we must note that dichotic stimulation showed no advantage over binaural stimulation on this measure. However, it is certainly far from being anything but a very gross measure.

How well do judges of pronunciation adequacy agree among themselves? In terms of rating scale values, we find that the degree of agreement is very low, the correlations ranging from .69 to .81. When total number of error words is used as the measure, the degree of agreement is a little higher, the coefficients ranging from .78 to .81. However, none of these values can be said to be very reassuring, especially

to a student being graded, at least in part, upon his ability to pronounce Spanish words and sentences.

We conclude, therefore, that dichotic stimulation has no advantage over binaural stimulation in the teaching of Spanish pronunciation as measured in this experiment. We recommend, in view of the marked differences and shortcomings in reliability and validity of judgments shown by our panels of experts in the counting of errors and the use of rating scale values, that objective tests of pronunciation adequacy be devised as soon as possible.

# BIBLIOGRAPHY


1. Borglum, George, "The Modern Language Audio-Visual Project," Modern Language Journal, XLII, 1958, 325-328

2. Calearo, C., "Binaural Summation in Lesions of the Temporal Lobe," Acta Laryngologie, XLVII, 1957, 392-395

3. Carhart, Raymond, "The Usefulness of the Binaural Hearing Aid," Journal Speech and Hearing Disorders, 23, 1958, 42-51

4. Cherry, E. C., "Some Experiments on the Recognition of Speech with One Ear and with Two Ears, "Journal Acoustical Society America, 25, 1953, 975-979

5. Johnston, Marjorie O., "The Urgency of Accelerating the Teaching of Foreign Languages," Modern Language Journal, XLII, 1958, 163-166

6. Johnston, Marjorie O., and Seerley, Catharine, "Foreign Language Laboratories in Schools and Colleges, Washington, D. C., United States Government Printing Office, 1958, Bulletin 1959, No. 3., 86 pp.

7. Koenig, W., "Subjective Effects in Binaural Hearing," Journal Acoustical Society America, 22, 1950, 61-62

8. Licklider, J. C. L., "The Influence of Inter-aural Phase Relations upon the Masking of Speech by White Noise," Journal Acoustical Society America, 20, 1948, 380-389

9. Ptaceck, Paul H., "An Experimental Investigation of Dichotic Word Presentation," 19, 1954, 412-422

10. Schubert, E. D., "Some Preliminary Experiments on Binaural Time Delay and Intelligibility," Journal Acoustical Society America, 28, 1956, 895-901

11. Van Riper, Charles, "Binaural Speech Therapy," Journal Speech Hearing Disorders, 24, 1959, 62-63

# APPENDIX A
# TRAINING TAPE MATERIAL

## Lección I

La instrucción

Monterrey

llueve

ingeniero

barrios

la civilización

?Cuál es la lengua nacional de México?

Los abreros viven en barrios pobres.

A lo largo de la costa la vegetación es abundante.

## Lección II

1. la cultura europea

México es un país de Norteamérica.

La capital de México es una ciudad muy grande.

Tiene casi tres millones de habitantes.

Es una ciudad bella.

Naturalmente, es el centro del gobierno mexicano.

La civilización de México es diferente de la civilización de los Estados Unidos.

Es una mezcla de la cultura india y de la cultura europea.

La mezcla de las dos culturas es interesante.

Laboratorio III

1. Monterrey
2. medio millón de habitantes
3. una ciudad industrial
4. Guadalajara
5. número
6. el mismo número de habitantes
7. Veracruz
8. puerto
9. Golfo de México
10. Tampico
11. exportación
12. petróleo

Hay otras ciudades de importancia de México.

La segunda ciudad es Monterrey.

?Cuántos habitantes tiene Monterrey?

Tiene casi medio millón de habitantes.

Es una ciudad industrial.

Recibe mucha influencia de los Estados Unidos.

La tercera ciudad de México is Guadalajara.

?Cúantos habitantes tiene Guadalajara?

Tiene el mismo número de habitantes que Monterrey.

Veracruz es un puerto ene el Golfo de México.

Tampico es otro puerto del Golfo de México.

Es importante por la exportación de petróleo.

Lección 4 (5-6)

1. montañoso
2. el pico más alto
3. el Orizaba
4. la meseta central
5. a causa de
6. agradable
7. tropical
8. llueve
9. vegetación

México es el país más montañoso del continente.
Cuál es el pico más alto de México?
El Orizaba es el pico más alto de México.
La capital se encuentra en la meseta central.
A causa de la altura el clima es agradable.
A lo largo de la costa el clima es tropical.
Llueve mucho y la vegetación es muy abundante.

Lección 5 (7)

1. la lengua nacional
2. pronuncian
3. exactamente
4. las diferencias de pronunciación
5. viven aislados
6. completamente diferentes

La lengua nacional de México es el español.
Los mexicanos hablan y pronuncian bien.
Sin embargo, no pronuncian exactamente como los españoles.
Las diferencias de pronunciación no son muchas.

Algunos indios viven aislados del resto del país.
Estos indios no hablan español.
Hablan dialectos indios completamente diferentes del español.
Existen casi cien dialectos diferentes.

Lección 6 (8-9)

1. la instrucción
2. no saben leer ni escribir
3. no hay ninguna
4. La Universidad Nacional

La instrucción en México es un gran problema.
Muchos mexicanos no saben leer ni escribir.
En algunas regiones no hay ninguna escuela.

Sin embargo, las escuelas de las ciudades son modernas.
Hay escuelas primarias, secundarias y profesionales.
Además, hay universidades en la capital y en otras ciudades.

La Universidad Nacional está en la capital.
Es la universidad más antigua del continente.

Lección 7 (9-10)

1. filosofía y letras
2. medicina
3. ingeniería

Roberto es norteamericano.
Es estudiante de la Universidad Nacional de México.
Estudia la historia y la cultura de México.

Felipe es un estudiante mexicano.
Estudia filosofía y letras en la Universidad.

Carlos es estudiante de medicina.

José estudia ingeniería.

# APPENDIX B
## TRAINING TAPE MATERIAL

(Second Period)

1. Buenos días. ?Cómo está usted?
   Muy bien, gracias, ?y usted?
   Muy bien. ?Está usted ocupado?
   No mucho.
   Vamos a dar un paseo.
   Muy bien, vamos.

2. ?Es usted norteamericano?
   Sí, señor, lo soy.
   Habla usted muy bien el español.
   Espero hablarlo mejor.

3. ?De dónde es usted?
   Soy de Medellin, Colombi.

4. ?Lleva usted mucho tiempo aquí?
   Unos cuantos días.
   ?Le gusta aquí?
   Si, señor, me gusta mucho.

5. Cuánto tiempo piensa usted quedarse?
   Voy a quedarme dos semanas más.
   ?Cuándo llegó usted?
   Llegué el día quince.

6. ?Ya se va?
   Sí. Tengo una cita.
   Pues, hasta luego.
   Hasta mañana.

7. Hola! ?Qu´e tal?
   No uy bien
   ?Qué le pasa?
   Estoy un poco cansado.

8. ?Ha viajado usted mucho?
   Un poco. Me queda mucho por ver.
   Todavía tiene usted tiempo.
   Sí, pienso viajar m-ás.

9. ?Le gusta au usted aquí?
   Muchísimo. Me gustar-ia vivir aquí.
   ?Y por qué no se quéda?
   Tengo qué volver a mi trabajo.

10. ?Ha visitado usted nuestro parque?
    Sí, estuve ayer.
    Es muy bonito.
    Sí, señor, me gustó much´ isimo.

11. Hace buen tiempo, ?verdad?
    Sí, hace un tiempo precioso.
    Siempre hace buen tiempo aquí.
    Me alegro.

12. ?Quién se lo dijo?
    Me lo dijo Pepe.
    ?Cuándo se lo dijo?
    Me lo dijo ayer.

13. Necesito diez dólares.
    ?Me los presta?
    No se los presto.
    ?Por qué no me los presta?
    Porque no los tengo.

14. ?Cuánto vale esto?
Diez pesos. Este otro vale ocho.
Muy bien. Me quedo con éste.
Se lo envuelvo, senor.

15. ?Puede usted cambiarme este billete?
Lo siento, pero no puedo. Tengo qu e pagar esta cuenta.
Se lo cambian en la caja.

16. Puede usted devolvérmelo?
no se lo puedo devolver hoy.
?Cuándo me lo puede devolver?
Se lo devuelvo manana.

17. Nos lo va a dar.
?Cuándo nos lo va a dar?
Nos lo va a dar mañana.
?No nos lo va a dar hoy?

18. Se lo he pedido prestado a Juan.
?Cuándo se lo ha pedido prestado?
Se lo pedí prestado ayer.
Estoy seguro de que se lo prestará.

19. ?Se le olvidó algo?
No, no se me olvidó nada.
Sí, senor, se le olvidó algo.
Ah sí, se me olvidó el dinero.

20. ?Ha perdido usted algo?
Sí, mi impermeable.
?Sabe usted dónde?
Probablemente, lo he dejado en el café.

21. ?Encontró su impermeable?
Sí, lo encontré.
?Lo dejó en el café?
No. Lo había dejado en el teatro.

## APPENDIX C

## TRAINING TAPES
## (Third Period)

1. Hace calor.
   Vamos a tomar un refresco.
   Bueno, vamos.
   ?Dónde iremos?

2. Dígamelo ahora.
   Se lo diré mañana.
   ?Por qué no me lo dice ahora?
   No puedo decírselo hasta mañana.

3. ?Qué va a tomar?
   Voy a pedir una coca-cola.
   Y yo tomo un café negro.
   Señorita, una coca cola y un caf'e negro, por favor.

4. Si quiere usted algo, pídamelo.
   Si quiero algo, se lo pediré.
   Puede pedírmelo ahora.
   Se lo pediré mañana.

5. ?Quiere usted tomar algo con nosotros?
   Sí, con mucho gusto.
   Siéntese usted. ?Qué quiere tomar?
   Voy a tomar un café con crema.

6. ?A quién se lo dará usted?
   Se lo daré a Juan.
   ?Cuándo se lo dará?
   Se lo daré mañana.

7. ?Cómo se llama esto en español?
   Eso se llama pluma fuente.
   ?Pluma fuente?...! Ah, lo mismo que en inglés.!
   Precisamente.

8. ?Cómo se dice <u>glasses</u> en espanol?
   Vasos.
   No. Para ver.
   !Ah! Se dice anteojos o gafas.

9. Esto se lo debo a Juan.
?No se lo debe a Pepe?
No, tampoco se lo debo a usted.
Pues, si se lo debe a Juan, déselo.

10. ?Qué piensa usted hacer esta noche?
No se. No he decidido.
?Le gustaría comer con nosotros?
Me gustaría mucho.

11. ?Quiere que se lo cuente?
Sí, cuéntemelo.
?Quiere que se lo cuente ahora?
Sí, quiero que me lo cuente ahora mismo.

12. ?Quiere usted ir al cine conmigo?
Con mucho gusto. ?Que ponen hoy?
Ponen una película muy buena.
Yo tomaré las entradas.

13. Tráigamelo usted.
?Quiere que se lo traiga?
Sí, quiero que me lo traiga.
Se lo traigo en seguida.

14. ?Cómo se llama este plato?
Es un plato típico del país.
¡Está riquísimo!
?No lo había probado antes?

15. ?Quiere usted que se lo repita?
Sí, repítamelo.
Se lo voy a repetir.
¡Está bien! No me lo repita más.

16. Parece usted muy contento.
Sí, estoy de buen humor.
?Por qué?
Acabo de recibir un cheque.
Hay que celebrar.

17. Búsquemelo usted.
Se lo voy a buscar más tarde.
Quiero que me lo busque ahora.
Muy bien, se lo buscaré ahora.

18. Tengo que marcharme.
Bueno, yo pagaré la cuenta.
Dejo la propina.
Como usted quiera.

19. Dice que me lo dé usted.
¿Quién dice que se lo dé?
Su padre dice que me lo dé.
Bueno, se lo doy.

20. ¿Cuánto debo dar de propina?
El diez por ciento o un poco más.
¿Es bastante esto?
Sí, creo que sí.

21. Explíquemelo.
¿Quiere que se lo explique?
¿No quiere explicármelo?
Sí, se lo voy a explicar.

22. Adiós, amigos. He pasado un buen rato.
Nosotros también. Hasta pronto.
Hasta la vista.
Que le vaya bien.

23. Me lo han prometido.
¿Quién se lo ha prometido?
Mis amigos me lo han prometido.
Me alegro de que selo hayan prometido.

24. Se lo mandaré mañana.
Mándemelo hoy,
No puedo mandárselo hoy.
Preferiría que me lo mandara hoy.

# APPENDIX D

## FIRST TEST TAPE

Words

1. la instrucción
2. Monterrey
3. llueve
4. ingeńiero
5. barrios
6. la civilizacioń

Sentences

1. Cuál es la lengua nacional de Mexico?
2. Los obreros viven en barrios pobres.
3. A lo largo de la costa la vegetacioń es abundante.

SECOND TEST TAPE

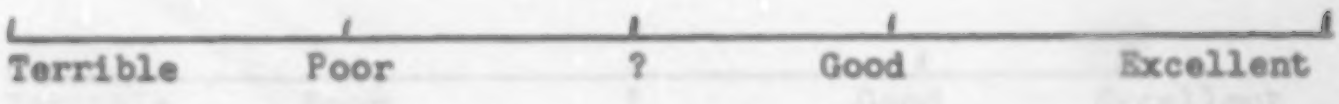
1. se lo cambian
2. devolvérmelo
3. nos lo va a dar
4. barrios
5. espero hablar lo
6. llegué
7. me gustó
8. le gusta

?Puede usted prestarme cinco dólares?

No puedo prestárselos hoy.

?Cuándo me los va a dar?

Voy a dárselos mañana.



THIRD TEST TAPE

1. bachillerato
2. Mediterráneo
3. el Guadalquivir
4. autorretrato
5. Córdoba
6. Alfonso XIII
7. álgebra
8. particular
9. amenazar
10. veterinario
11. Navarra
12. generalísimo

Dígamelo.

Se lo diré el lunes.

?Por qué no me lo dice ahora?

No quiero decírselo hasta mañana.

?Quiere usted tomar algo con nosotros?

Bien, voy a tomar un café negro.

Señorita, un café negro y una coca cola, por favor.

Esto se lo debo a Felipe.

Pues si se lo debe a Felipe, déselo.

Terrible Poor ? Good Excellent